Admixture between rarely hybridizing milkweeds, *Asclepias syriaca* and *Asclepias sullivantii*: Morphological and molecular evidence from populations across the Midwestern U.S.

Undergraduate Research Thesis

Presented in Partial Fulfillment of the Requirements for graduation "with Research Distinction" in the undergraduate colleges of The Ohio State University

by

Nathaniel P. Hofford

The Ohio State University

April 2020

Project Advisors: Dr. Stephen M. Hovick & Dr. Andreas S. Chavez

Abstract

Interspecific hybridization and subsequent admixture often has important ecological and evolutionary consequences, especially when one of the species involved is relatively rare. Common milkweed (*Asclepias syriaca*) has been observed to hybridize with its uncommon congener *A. sullivantii*, but the extent to which hybridization occurs across their range is unknown. The purpose of this study was to quantify the extent of admixture between *A. syriaca* and *A. sullivantii* in multiple co-occurring populations across the Midwest U.S. using morphological and genomic data. We collected morphological leaf data and performed ddRADSeq on 276 *Asclepias* individuals. Morphological analyses indicate distinct parental phenotypes, and moderate phenotypic deviation from parental taxa by putative hybrids. Preliminary STRUCTURE plots indicate extensive admixture in most co-occurring populations, although definitive conclusions cannot be made based on a low burn-in and number of MCMC repetitions for these analyses. Our work represents an important follow up of a previous study into this system, expanding the geographic area sampled and providing high resolution genomic data.

Introduction

Hybridization occurs when two parents of different species produce viable offspring. Hybridizing systems offer a unique perspective into the evolutionary effects of gene flow between species. Even if hybridization events are rare and the fitness of F1 individuals is low, this process can still have beneficial effects on advanced generation taxa due to the infusion of novel genes into a population (Allendorf et al., 2001; Arnold, 1997). Conversely, hybridization can result in genetic swamping and production of hybrid swarms when either prezygotic or postzygotic isolating barriers are weak. In this case, genetic identity of one or both parental species can be lost, which can be especially problematic for hybridization between rare and common taxa (Rhymer & Simberloff, 1996).

Despite several examples of hybridizing systems, concurrent flowering times, and shared pollinators, hybridization events among milkweed species (in the genus *Asclepias*) are relatively rare (Kephart & Theiss, 2003.; Klips & Culley, 2004; Wyatt & Broyles, 1994; Woodson, 1954). Evidence suggests that physiological and mechanical barriers, such as pollen tube formation and size differences between stigmatic chamber and pollinia (coherent mass of pollen grains), could be driving hybrid rarity (Hatfield & Kephart, 2003). Furthermore, there may be a paradoxical relationship between the frequency of hybridization and degree of introgression in *Asclepias* systems. Studies of hybridization between *A. exaltata* and *A. syriaca* found a high degree of later-generation backcrossing from between hybrids and pure parental lines and low frequency of F1 generation hybrids (Broyles, 2002; Broyles et al., 1996). It is unclear if this same pattern occurs in other hybridizing *Asclepias* systems.

We chose to examine two *Asclepias* species, *A. syriaca* and *A. sullivantii*, which co-occur across much of the Midwest. *A. syriaca*, otherwise known as common milkweed, has a wide ecological and geographic range (Cooperrider, 1997; Gleason & Cronquist, 1991). In the United States, its geographic range spans the Midwest and east coast, with non-native populations in Oregon (Kartesz, 2015). The range of *A. sullivantii* (Sullivant's milkweed) is much more limited, and populations are restricted to intact prairies ranging latitudinally from central Minnesota to Oklahoma and longitudinally from eastern Ohio to central Nebraska (Kartesz, 2015).

Hybridization between *Asclepias syriaca* and *Asclepias sullivantii* was first studied in Klips & Culley, 2004. They confirmed natural hybridization at a single Ohio site (Killdeer Plains) via four main lines of evidence: flowering time overlap, morphological differences, isozyme markers, and experimental hand pollinations. Their study confirmed these species' hybridization potential by establishing overlapping flowering times, distinct morphological differences between parents and hybrids, and more frequent hybrid heterozygosity (at tested loci). Importantly, the study also determined via hand pollinations that hybrid fruit could be set when *A. syriaca* received pollinia from *A. sullivantii*, but no fruit was set when *A.*

sullivantii acted as the ovule parent. Pollinations also showed that *A. syriaca* individuals could set fruit when receiving a pollinia from hybrid individuals, although at a much lower frequency.

Our investigation into the same system re-examines the morphological and genetic component of the Klips & Culley study on a broader scale, sampling from populations across the Midwest United States. Additionally, we employed a genomic sequencing technique, double digest restriction site associate sequencing (ddRADSeq), which generates fragments across the nuclear genome for population genetic inference (Peterson et al., 2012). Utilizing ddRADSeq allows us to assess thousands of loci per population, as opposed to the three used in the Klips & Culley study. The aim of our study was to test for the morphological and genomic signature of hybridization across the species co-occurring range.

Methods

Sample Collection

Our sampling effort consisted of collecting from 27 total *Asclepias* populations across the Midwest, including Ohio (OH), Illinois (IL), Wisconsin (WI), and Iowa (IA). At each site, 10-15 putatively distinct individuals of each species were collected on the basis of morphology. Hybrids were collected and labeled as such if they could be identified in the field. These field identifications served as the basis for our labeling of an individual as either *A. syriaca*, *A. sullivantii* or putative hybrid throughout the study. One mid-stem leaf was collected and pressed for morphological analysis and three to five young top-stem leaves were collected for DNA extractions. These young leaves were stored in a cooler over ice until they were transferred to a -80 °C freezer before extraction.

Initially, over 550 plants were sampled, but because of DNA extraction issues and cost constraints, our final data set consists of 276 *Asclepias* individuals (188 *A. syriaca*, 84 *A. sullivantii*, and 4 putative hybrids) representing 18 populations. In this final subsampling, each population contained at least five individuals from each species that occurred there. Of the 18 populations, nine are *A. syriaca* only

populations, and nine have co-occurring *A. syriaca* and *A. sullivantii*. All nine *A. syriaca* only sites are from OH, whereas co-occurring sites included two from IA, three from WI, one from IL, and three from OH (Figure 1).

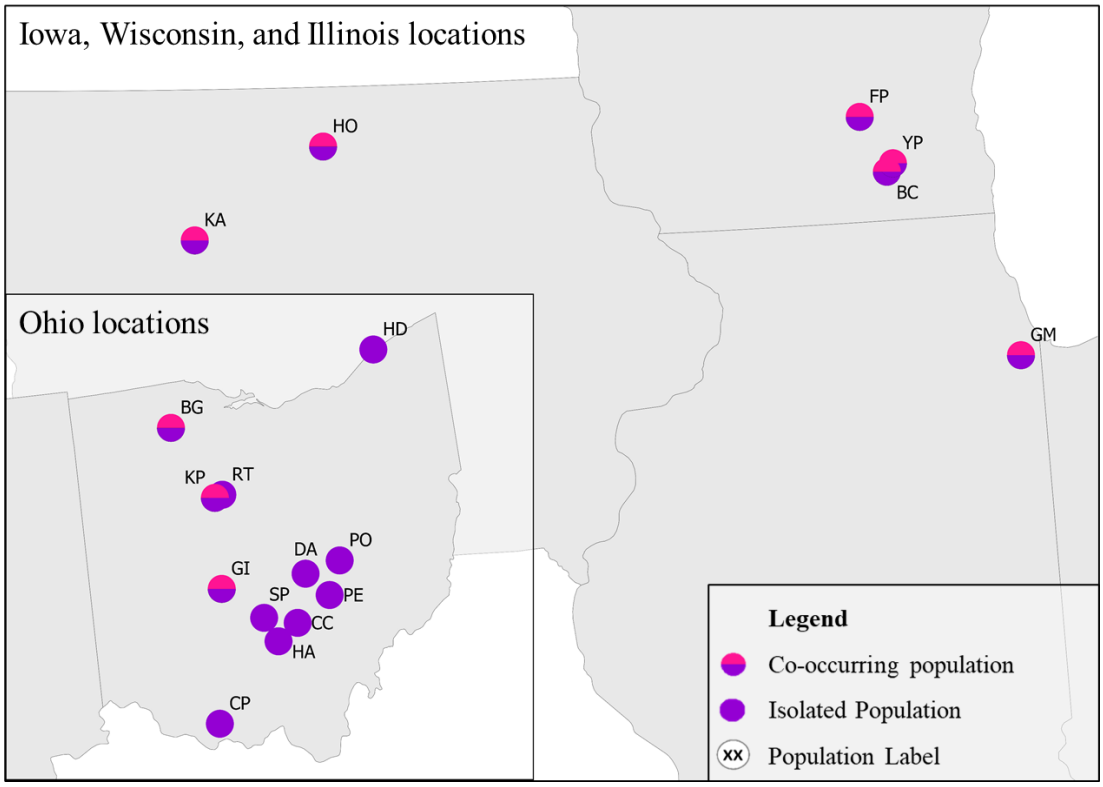

Figure 1: Map of sampling location across the Midwest. Isolated (*A. syriaca* only) populations are represented by purple points and occur only in Ohio (inset map). Co-occurring populations (*A. syriaca* and *A. sullivantii*) are represented by pink and purple points (both maps).

Morphological Data Collection

To quantify variation in leaf morphology between putative species, a measurement protocol was adopted from Klips & Culley, 2004. A total of seven leaf characteristics were used, including one which was a derived ratio. Petiole length was quantified from the base of the leaf the tip of the petiole. Leaf length was measured from the base of the blade to the tip, and leaf width was measured at half the leaf length. Three angles were measured to better capture overall leaf shape. Angle A measures the ovateness of the leaf,

angle B measures the cordateness of the base, and angle C measures the acuteness of the apex. Finally, a ratio was taken between the leaf width and length (Figure 2).

Pressed and dried leaves were first scanned with an Epson Perfection V800 scanner (Epson America Inc., San Jose, CA, USA). Using the image analysis program WinFolia (Regent Instruments Inc., Canada) we created measurement guidelines. These scans were then transferred to Adobe Photoshop where measurements and angles were quantified using the ruler tool.

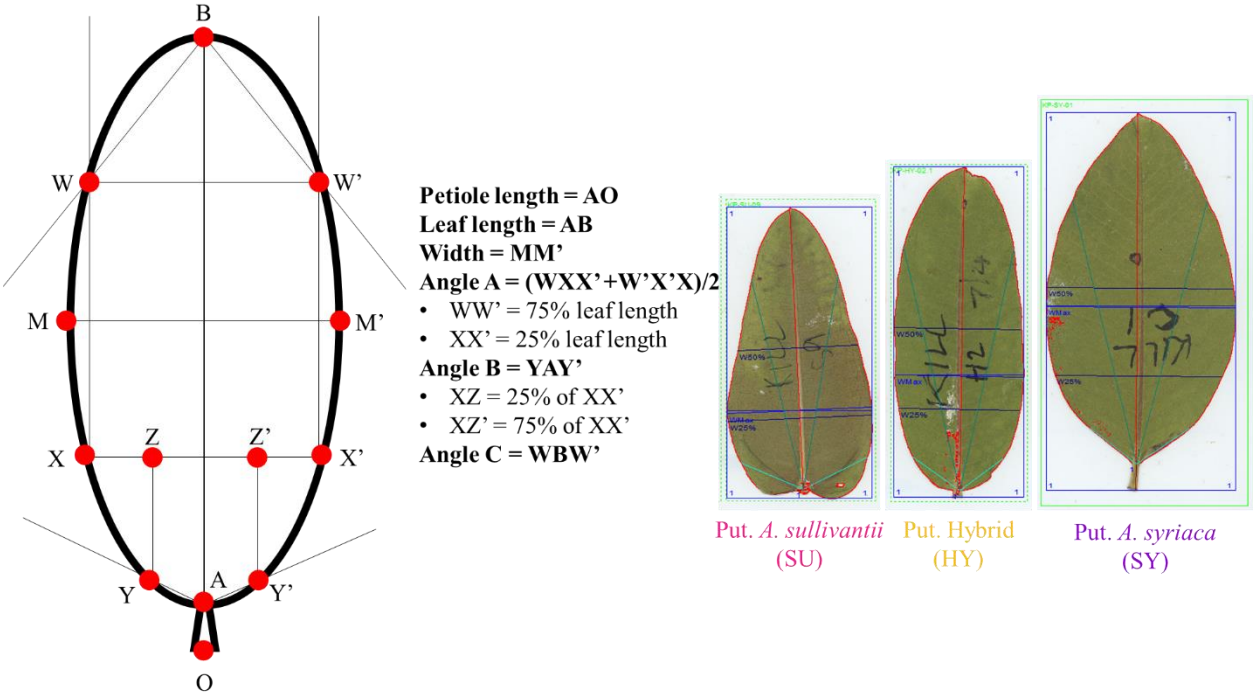

Figure 2: *Left*, Outline of leaf measurement protocol, redrawn from Klips & Culley, 2004. Letters depict points used in measurements. *Right*, Leaf samples with measurement guidelines drawn using WinFolia from each putative group (SU, HY, and SY) collected from Killdeer Plains (KP) in Ohio.

Morphological Analyses – Kruskal Wallis Test & Principal Component Analysis

Summary statistics for all 276 samples, split by putative species designations, were collected using the R package dplyr (Wickham et al., 2019). Data were visualized using the ggplot2 package in R (Wickham, 2016). Initial One-Way ANOVA tests were computed for each trait to compare putative species, and the

residuals were plotted against the fitted model values. A Levene's test in the car R package was used to assess the statistical significance of the residual's variance (Fox & Weisberg, 2019). An additional non-parametric analysis, the Kruskal-Wallis test, was performed on the data once Levene's test and residual plots revealed the data to have unequal variance. Pairwise comparisons between species for each trait were further assessed using a Bonferroni corrected Wilcoxon Rank Sum test.

To more holistically assess the morphological variation between putative species, a principal component analysis was carried out with all seven characteristics. Using the GGally extension of ggplot2 in R, a correlation matrix of Pearson's Product-Momentum Correlation values (r) was first computed to assess if any highly correlated traits should be removed from the analysis (Schloerke et al., 2020). All trait correlations fell under the $r > 0.90$ cutoff, so all were included in the final analysis. Carried out in the vegan package of R, trait data were scaled during the PCA to account for variation in their ranges (Oksanen et al., 2019).

DNA Extraction

A modified cetyl trimethylammonium bromide (CTAB) protocol optimized for high-throughput was used to extract DNA from the collected young leaf tissue. Approximately 0.2 g of frozen tissues sample was homogenized using dry ice and homogenizer beads, broken up at a frequency of 25 Hz for two minutes in 1.1 mL tubes. Samples were extracted with a 3X CTAB solution containing 0.03 g/mL of PVP and 30 µL/mL of betamercaptoethanol heated for 10 minutes in a 65˚C water bath. 400 µL of the CTAB solution was transferred to each sample tube and homogenized for two minutes at 25 Hz. To remove polysaccharide contamination, samples were incubated for 30 minutes in a 65˚C water bath and centrifuged for 30 minutes at 3,500 rpm. The aqueous layer at the top of the tube was transferred to new 1.1 mL tubes where 400 µL of 24:1 chloroform:isoamyl alcohol was added to remove additional secondary contaminants and isolate DNA. Tubes were inverted 50 times and left to rest on their side (to increase the surface area between the choloroform and CTAB layers) for five minutes. These tubes were then centrifuged for one hour at 3,500 rpm after which the top aqueous layer was transferred to new tubes.

400 μL of cold (-20˚C) 100% isopropanol was added to each tube that were then inverted 50 times before being stored at 4˚C overnight. Post storage samples were centrifuged for 30 minutes at 3,500 rpm. The supernatant was decanted, leaving a small pellet at the bottom of the tube. This pellet was air dried before 200 μL of water was added, and gentle pipetting broke the pellet up until it went back into solution. 400 μL of cold (-20˚C) 100% ethanol was added to each tube before being inverted 50 times and stored in -20˚C for 30 minutes. Samples were then centrifuged for 30 minutes at 3,500 rpm and again decanted and the pellet dried. For final resuspension and storage, 101 μL of a TE+RNase (100 μL of TE buffer, 1 μL of RNase) solution was added to each dried pellet and allowed to sit at room temperature for one hour. DNA concentration and contamination ratios were measured on the NanoDrop One Microvolume UV-Vis Spectrophotometer (Thermo Scientific NanoDrop Products). Samples with a concentration ≥ 150 ug/μL and 260/280 and 260/280 ratios ≥ 1.8 were cleared for the ddRADSeq protocol, as low concentration and contaminated DNA would likely interfere with the restriction enzymes in the ddRADSeq procedure (Graham et al., 2015).

ddRADSeq

In order to generate genome-wide SNPs, we used a double-digest restriction-site associated sequencing (ddRADSeq) protocol following the guidelines of Peterson et al. (2012) and slightly modified to accommodate our specific system. Samples were first diluted in 0.3 μL PCR tubes to 162.8 ng/μL for a total of 7000 ng of DNA in 43 μL. Such a high amount of starting DNA was used to account for possible inaccuracies in the Nanodrop concentration readings.

Samples were digested with two restriction enzymes (RE), MspI and SpfI-HF (New England Biolabs). 7 μL of the RE master mix (containing 1 μL of each RE and 5 μL of CutSmart buffer) was added to the 43 μL samples. Samples were digested at 37˚C for two hours followed by a 4˚C hold. A 1.5X Serapure bead (GE Healthcare Sera-Mag SpeedBeads™ Carboxyl Magnetic Beads, Thermo Fisher Scientific) cleanup at 75 μL was performed and 30 μL supernatant was transferred to new 0.3 mL tubes.

Before ligation of DNA barcodes, an in-silico digest of the *A. syriaca* genome was performed using the rsitesearch program in the Python package ddradseqtools (Mora-Márquez et al., 2017). This was used to refine the ligation molarity calculator by providing more accurate estimates of the RE cut site frequencies. T4 ligase was combined with 40 μM P2 adapter and eight unique P1 adapters. Across each strip tube, 7 μL of a unique adapter was added to the DNA sample. Samples were then incubated on a thermocycler at 37˚C for 30 minutes, 65˚C for 10 minutes, and cooled by 2˚C increments per minute until reaching 10˚C. 12 eight-sample strip tubes were pooled to a total volume of 320 μL (40 μL per sample). Two 1X bead washes were carried out, the first eluting to 50 μL and the second to 30 μL. Samples were run on the Agilent 2200 Tape Station (D100 ScreenTape, Agilent Technologies, Walderbronn, GE) to check for 90 bp dimers before size selection. Fragments ranging from 250 – 600 bp were selected based on the in-silico digest following the guidelines of the BluePippin manual (Sage Science, Beverly, MA).

At the final step before sequencing, Illumina indexes were annealed, and PCR was carried out to increase the total amount of DNA in each pool. 25 μL 2X Phusion master mix, 2 μL of 12 unique P5 and P7 indexes, and 21 μL of pooled DNA were added to new 0.2 mL strip tubes. Samples were placed in a thermocycler and incubated for 30 seconds at 98˚C, followed by 18 cycles of 98˚C for 10 seconds, 65˚C for 30 seconds, and 72˚C for 30 seconds. Samples were held at 72˚C for five minutes before a 10˚C final hold. All pools then underwent a 1.5X Serapure bead cleanup (75 μL) and run on the Agilent tape station to check for adapter dimers. Pools with dimers received an additional three bead washes at 1.5X (75 μL), 1.25X (62.5 μL), and 1.15X (57.5 μL), checking pools on the Agilent tape station in between each wash.

Final libraries were pooled at a molarity of 3 nM in 20 μL of diH_2O. Libraries were sequenced at the Genomics Shared Resource Center at The Ohio State University (Genomics Shared Resource, The Ohio State University Comprehensive Cancer Center, Columbus, OH) on the Illumina Hi-Seq 4000 (Illumina, Inc.), yielding 150 bp paired-end reads.

Genomic Assembly and Analyses – iPyRAD and STRUCTURE

For the preliminary genomic analysis, only the co-occurring populations were analyzed (sites where both *A. syriaca* and *A. sullivantii* occur). 179 total samples were assessed, comprising nine total populations and consisting of four putative hybrids, 84 putative *A. sullivantii*, and 91 putative *A. syriaca*.

Assembly of raw paired-end reads was completed via the *de novo* method in the program iPyRAD version 0.9.20 (Eaton & Overcast, 2020). After quality filtering, R1 reads were clustered at an 85% threshold within each sample. Individual sample heterozygosity and error rate were jointly estimated through counts of site patterns on clustered reads and used for consensus base calling. Average heterozygosity and sum of squares error (SSE) for a putative species at a population were calculated based on these values. A pairwise t-test using the average heterozygosities at each population as the unit of replication (n = 9) was carried out assess if species-specific averages differed significantly. Consensus loci were then aligned across samples at 85% similarity. At each population, loci were filtered so they were shared by at least four individuals before single nucleotide polymorphism (SNP) matrices were generated for each population.

We used the population clustering program STRUCTURE version 2.3.4 to assess potential admixture within each population (Pritchard et al., 2000). A burn-in of 10,000 with 100,000 Markov chain Monte Carlo repetitions was used for each run of STRUCTURE. We used the default STRUCTURE parameters, the admixture model, and three iterations of K=2 population size. Outfiles were concatenated using the Clustering Markov Packager Across K (CLUMPAK) (Kopelman et al., 2015). Final plots were built using the program STRUCTURE PLOT (Ramasamy et al., 2014). All analyses were carried out using the Unity Supercomputer Cluster (The Ohio State University, College of Arts and Science Technology Service, Columbus, OH).

Results

Morphological Data

Residual plots and Levene's test indicated that ANOVA was not appropriate to assess differences between species for each trait, and instead a non-parametric Kruskal-Wallis test had to be performed (Table 1). Residual plots showed that variation in the putative hybrid group was not equivalent to the parental taxa (Figure S1). Only one trait had a non-significant p-value for Levene's test, Angle A, however, it was still analyzed using a Kruskal-Wallis test after examining the non-normal residual plot. Kruskal-Wallis tests identified significant differences among groups in all seven leaf traits (Table 1).

Character Trait	ANOVA		Levene Test		Kruskal-Wallis Test	
	p-value	sig-code	p-value	sig-code	p-value	sig-code
Petiole Length, p (cm)	2.00E-16	***	3.22E-09	***	2.20E-16	***
Leaf Length, q (cm)	1.77E-09	***	2.54E-04	***	1.00E-09	***
Leaf Width, r (cm)	3.22E-16	***	1.98E-03	**	2.51E-16	***
Angle A, s (degrees)	1.07E-14	***	8.50E-01	NS	3.35E-13	***
Angle B, t (degrees)	2.00E-16	***	4.84E-02	*	2.20E-16	***
Angle C, u (degrees)	1.41E-13	***	4.09E-02	*	7.67E-15	***
Ratio of width to length, r:q	2.31E-08	***	5.12E-02	.	1.71E-07	***

Table 1: Summary p-values from ANOVA, Levene's test, and the Kruskal-Wallis test for each leaf trait. Displayed here are p-values and significance codes ($p < 0.000$ '***' ; $p < 0.001$ '**' ; $p < 0.010$ '*' ; $p < 0.050$ '.' ; $p > 0.050$ 'NS").

Pairwise, Bonferroni corrected Wilcoxon rank sum tests indicated that putative parental groups differed significantly from each other in all seven leaf traits (Table 2, Figure 3). Putative species means for various

traits indicated that on average, *A. syriaca* individuals had longer petioles, were more ovate (larger Angle A), and were wider and longer than *A. sullivantii* individuals. Angle B, measuring the cordateness of the leaf base was much larger in *A. sullivantii* individuals than in *A. syriaca*. Finally, *A. syriaca* individuals had less acute leaf blades than *A. sullivantii* (Angle C).

Putative hybrid taxa were never statistically distinct from both parents but did differ from a single parent at specific traits. Putative hybrids differed significantly from *A. syriaca* with regards to petiole length, having a smaller average length than the parent. When compared to *A. sullivantii* taxa, hybrid mean trait values significantly differed for leaf width, Angle C, and the width:length ratio. At all three of these traits, hybrid mean values exceeded both parental taxa means, having wider leaves, less acute leaf blades, and a higher width:length ratio.

Character Trait	Put. Hybrid (n = 4)	Put. *A. sullivantii* (n = 84)	Put. *A. syriaca* (n = 188)
Petiole Length, p (cm)	0.491 ± 0.0968^{A}	0.330 ± 0.133^{A}	0.953 ± 0.319^{B}
Leaf Length, q (cm)	14.4 ± 2.06^{AB}	12.7 ± 1.95^{A}	15.1 ± 3.05^{B}
Leaf Width, r (cm)	6.43 ± 0.475^{A}	4.52 ± 1.10^{B}	6.22 ± 1.58^{A}
Angle A, s (degrees)	86.3 ± 2.58^{AB}	83.8 ± 2.87^{A}	87.0 ± 2.91^{B}
Angle B, t (degrees)	153 ± 26.3^{AB}	164 ± 32.2^{A}	116 ± 24.5^{B}
Angle C, u (degrees)	73.2 ± 9.38^{A}	57.1 ± 8.32^{B}	68.1 ± 11.3^{A}
Ratio of width to length, r:q	0.454 ± 0.068^{A}	0.354 ± 0.067^{B}	0.415 ± 0.082^{A}

Table 2: Mean and standard deviation subdivided by putative group and trait. Significant pairwise comparisons from the Bonferroni corrected Wilcoxon rank sum test displayed as superscript.

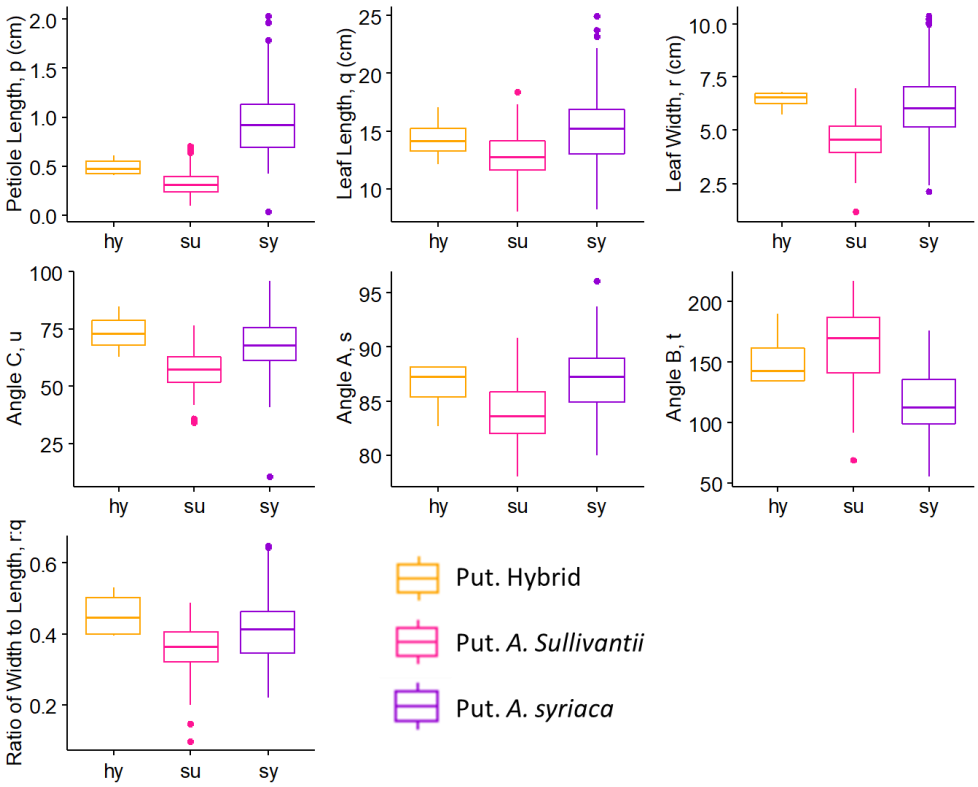

Figure 3: Box plots of leaf morphological characteristics. Putative hybrid samples are from KP (the only site with put. hybrids) while *A. syriaca* and *A. sullivantii* are aggregated across all 18 populations.

The correlation matrix revealed moderately strong correlations (0.60 < r < 0.80) between width and length, width and width:length, and angle A and angle B (Figure S2). Additionally, between angle C and width:length there was a stronger correlation of r = 0.859. For the plotted PCA, PC1 explained 41.93% of the variation and PC2 explained 29.65% (Figure 4). PC1 was primarily associated with width, angle C, and length. PC2 was primarily associated with angle B and angle A. While there is overlap between the three taxa, some separation can be seen occurring on PC1 and along the diagonal of quadrant III, associated with angle B. The putative hybrid group overlaps with *A. syriaca* but is positioned intermediate to the *A. sullivantii* individuals.

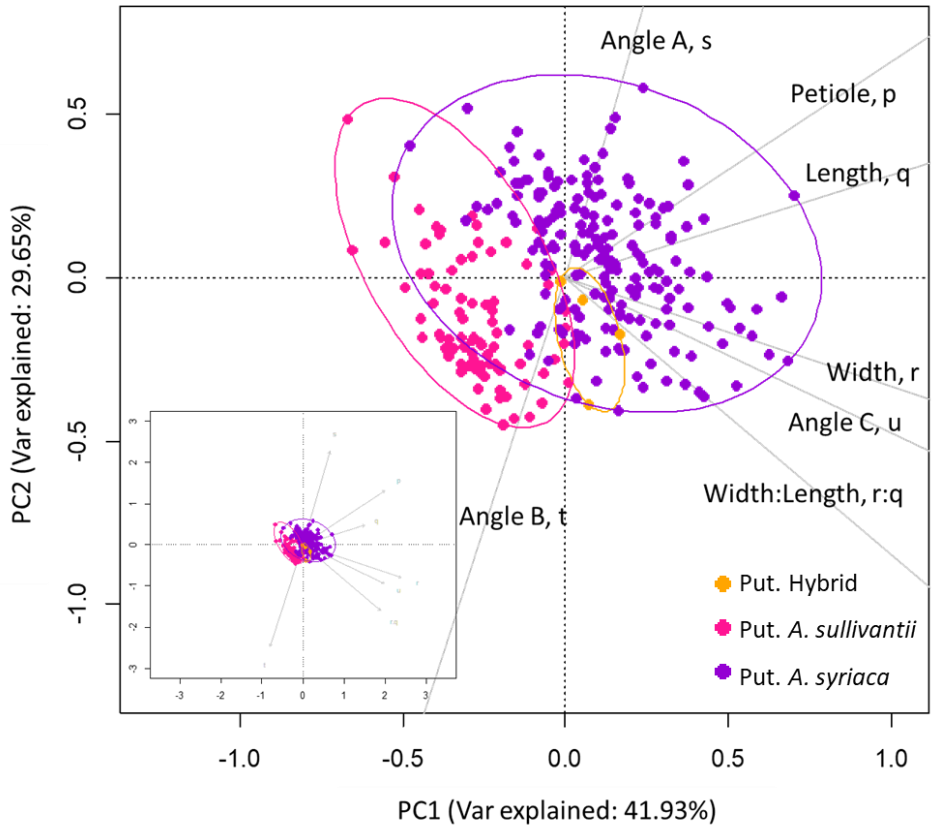

Figure 4: Zoomed PCA of full leaf trait data, colored by putative species. Inset PCA shows zoomed relative contribution of each loading.

Genomic Data

iPyRAD *de novo* assembly compiled an average of 2,875,567 ± 3,495,838 raw sequence reads per sample. After processing, an average of 7559 ± 4134.72 loci were recovered per population (Table 3). Average SNP matrix size for each population was 30001 ± 21200.95 with missing data per matrix averaging 58.04% ± 7.739. Across all populations, *A. syriaca* heterozygosity was 0.0202 ± 0.0306 and *A. sullivantii* heterozygosity was 0.0184 ± 0.0242. Only the Killdeer Plains population contained putative hybrids, whose average heterozygosity was 0.0223 ± 0.0040. Comparison of the parental taxa heterozygosities with a pairwise t-test was not significant ($p = 0.13$).

Population (ID)	State	Total Individuals (n=179)	Put. A. syriaca (n=91)	Put. A. sullivantii (n=84)	Put. Hybrids (n=4)	Total Loci	SNPs	Missing Data	Average heterozygosity HY	Standard error HY	Average heterozygosity SU	Standard error SU	Average heterozygosity SY	Standard error SY
Bluff Creek (BC)	WI	12	5	7	0	4243	10366	49.34	NA	NA	0.0207	0.0056	0.0205	0.0051
Bowling Green (BG)	OH	14	6	8	0	3549	9274	50.30	NA	NA	0.0193	0.0076	0.0185	0.0048
Faville Prairie (FP)	WI	25	13	12	0	3729	15441	55.40	NA	NA	0.0165	0.0092	0.0195	0.0036
Gillivan (GI)	OH	17	11	6	0	12242	34655	64.12	NA	NA	0.0180	0.0069	0.0205	0.0043
Gensburg-Markham (GM)	IL	20	14	6	0	10410	44408	65.16	NA	NA	0.0169	0.0064	0.0165	0.0040
Hoffman Prairie (HO)	IA	21	11	10	0	1598	6545	44.03	NA	NA	0.0193	0.0071	0.0193	0.0039
Kalsow Prairie (KA)	IA	17	8	9	0	10544	53910	60.65	NA	NA	0.0185	0.0077	0.0234	0.0043
Killdeer Plains (KP)	OH	33	17	12	4	14917	74998	69.51	0.0233	0.0040	0.0196	0.0096	0.0246	0.0031
Young Prairie (YP)	WI	20	6	14	0	6803	20415	63.87	NA	NA	0.0171	0.0112	0.0158	0.0040
Average	NA	20	10	9	NA	7559	30001	58.04	NA	NA	0.0184	0.0242	0.0202	0.0306
Standard deviation	NA	5.59	3.70	2.57	NA	4134.72	21200.95	7.739	NA	NA	NA	NA	NA	NA

Table 3: iPyRAD output data for each population assembly. Number of individuals, loci, SNPs, missing data, heterozygosity, and SE are all reported. Note that species specific average heterozygosities reported here were calculated using the individual as the unit of replication, without regard to population assignments.

Preliminary STRUCTURE results (Figures. 7-13) are difficult to interpret definitively due to a low burn-in and relatively small number of MCMC reps. Columns represent individuals and the proportion of either purple (*A. syriaca*) or pink (*A. sullivantii*) colors within each bar represents the relative population assignment of that sample. Pure taxa are shown as monochromatic columns, while individuals belonging to multiple populations (i.e. putative admixture) are represented by bars containing both colors. While most individuals segregated according to their putative species IDs, several samples appear to be mis-identified, either in the field or by STRUCTURE. Faville Prairie (FP) contains one such individual in the putative *A. syriaca* group that, according to STRUCTURE, falls completely in the *A. sullivantii* population (Figure 7). In almost all populations (FP, GI, GM, HO, KA, KP and YP) at least one individual contains > 0.75 membership coefficient from the opposite species that it was morphologically described as belonging to (Figures 7-13). Possible admixture occurs in all populations except Bowling Green (BG) (Figure 6).

Putative hybrid individuals at Killdeer Plains had variable population assignment (Figure 12). Two contain >0.75 membership coefficients for the *A. sullivantii* population, and another >0.75 membership coefficient for the *A. syriaca* population. Finally, one individual fell completely into the *A. syriaca* population with a membership coefficient = 1.

Figure 5: STRUCTURE output for Bluff Creek (BC) population. Population size = 12; Put. SY = 5; Put SU = 7.

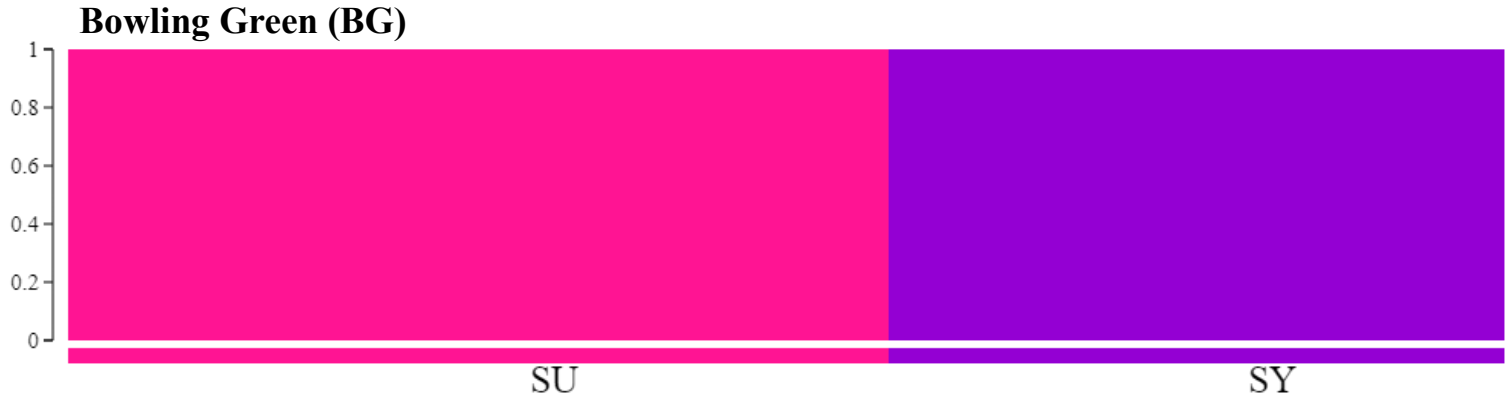

Figure 6: STRUCTURE output for Bowling Green (BG) population. Population size = 14; Put. SY = 6; Put SU = 8.

Figure 7: STRUCTURE output for Faville Prairie (FP) population. Population size = 25; Put. SY = 13; Put SU = 12.

Figure 8: STRUCTURE output for Gillivan Prairie (GI) population. Population size = 17; Put. SY = 11; Put SU = 6.

Figure 9: STRUCTURE output for Gensburg Markham (GM) population. Population size = 20; Put. SY = 14; Put SU = 6.

Figure 10: STRUCTURE output for Hoffman Prairie (HO) population. Population size = 21; Put. SY = 11; Put SU = 10.

Figure 11: STRUCTURE output for Kalsow Prairie (KA) population. Population size = 17; Put. SY = 8; Put SU = 9.

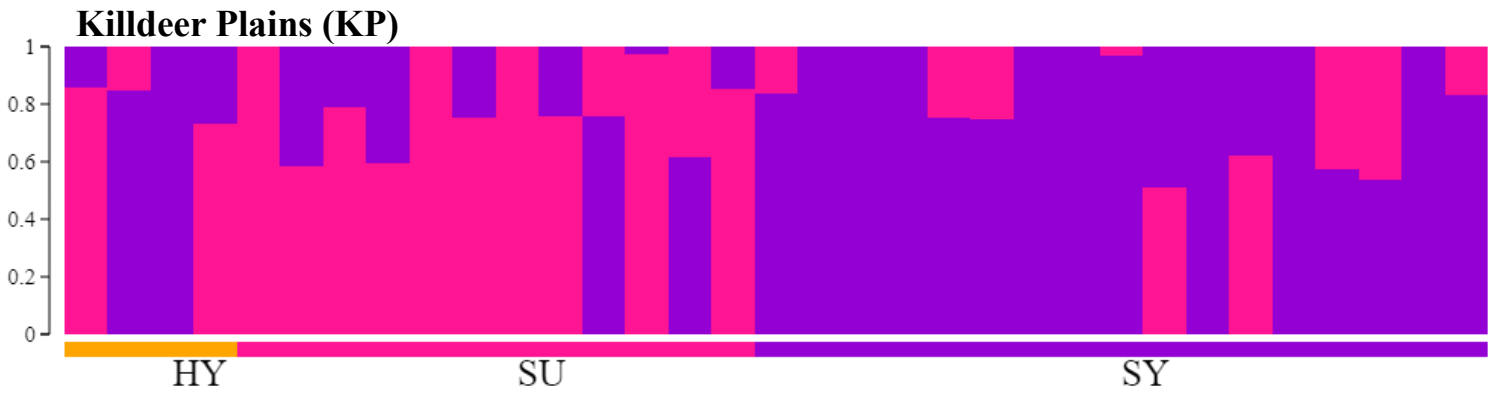

Figure 12: STRUCTURE output for Killdeer Plains (KP) population. Population size = 33; Put. SY = 17; Put SU = 127; Put. HY = 4.

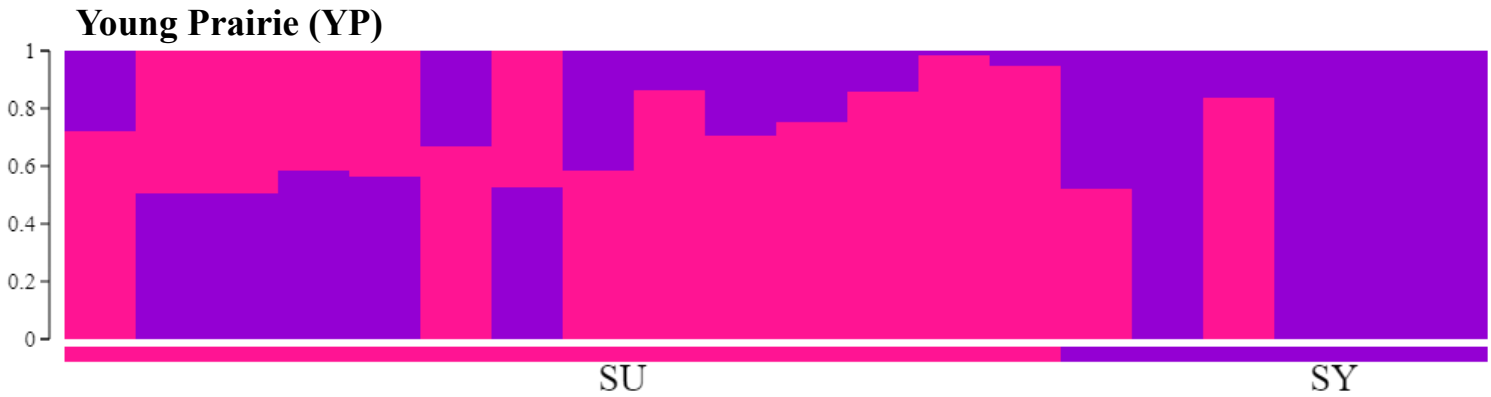

Figure 13: STRUCTURE output for Young Prairie (YP) population. Population size = 20; Put. SY = 6; Put SU = 14.

Discussion

The data collected through this study represents a novel and more comprehensive sampling of a previously studied hybridizing species pair. Compared to the only other investigation into this system (Klips & Culley, 2004), our study broadens the geographic range and increases the amount of genetic data collected for these species. Morphological evidence demonstrates that parental species are distinct, and that putative hybrid individuals differ at specific vegetative traits. Genomic analyses via STRUCTURE show possibly extensive admixture between putative parental populations, but definitive interpretation is limited due to low burn-in and MCMC repetitions.

Our morphological data largely mirrored the leaf trait results of Klips & Culley, who also included pollinia and flower measurements in their analysis. Both studies found significant differences between parental taxa at all seven leaf traits, and that petiole length and angle B are useful metrics for distinguishing taxa. We confirmed the Klips & Culley result that hybrids never differed significantly from both parents at any vegetative traits but found different significant comparisons when looking at a single parent and the hybrid. Klips & Culley found hybrids had significantly shorter petioles, a smaller width, and a lower width:length ratio than *A. syriaca*, in addition to being intermediate to the parents at every trait. Our study showed that petioles of the putative hybrids were significantly shorter than those of *A. syriaca*, but that width and width:length were significantly different than *A. sullivantii*, and larger than both parental mean values. We also identified angle C as significantly larger in hybrids than in *A. sullivantii*. These were the only non-intermediate hybrid values for our data set, whereas Klips & Culley found all hybrid leaf traits were intermediate. Klips & Culley also showed that angle B was significantly larger in hybrids than *A. sullivantii*, but we found that the hybrids were not significantly different than either parent for this trait. Because our sampling was widespread throughout the Midwest, the disparity between these data could reflect a broader vegetative trait space for the two parental taxa. One possible artefact for hybrid comparisons could also be our low sample size of putative hybrids, although the Klips & Culley study only sampled seven total individuals (compared to our four). Another possibility is the

misassignment of individuals to putative groups. Our analyses were based on field identifications of individuals, which could have failed to capture an important morphological gradient caused by admixed individuals.

It is important to keep in mind that the STRUCTURE runs completed for this study are difficult to interpret definitively due to a low burn-in and MCMC repetitions. However, if the extensive pattern of admixture indicated at the majority of co-occurring sites holds with further analysis, gene flow between these species is much more extensive than previously estimated in Klips & Culley, 2004. Comparisons between the Klips & Culley isozyme data, and the genomic data we collected are limited due to the difference in scale and data type. Klips & Culley collected data for three shared loci across all individuals, whereas our data encompasses thousands of loci that are not shared by all individuals in a population (i.e. the missing data in our SNP matrices).

Further comparisons can be drawn with patterns of gene flow in other hybridizing *Asclepias* systems. Previous studies investigating the relative frequency of F1 hybrid formation vs. introgression between *A. syriaca* and *A. exaltata* have shown that hybrid formation is rare, but stable hybrids in the population are able to backcross with both parental tax (Broyles, 2002). Broyles showed that established hybrids receive pollinia and mate with parents at a greater frequency than parents mate with each other, and thus the formation of advanced generation hybrids is accelerated over F1 formation. While we cannot sort putatively admixed individuals into distinct classes yet, our data will enable us to test this hypothesis by assessing the proportion of F1, advanced generation hybrids, and backcrossed individuals empirically.

Further investigation into this system is necessary for a variety of reasons. There are conservation implications for *A. sullivantii*, a relatively rare species, if frequent admixture is occurring at all sites where it co-occurs with *A. syriaca*. Rarer taxa can become genetically swamped out by the more common species (Ellstrand, 1992), and thus admixture could potentially affect *A. sullivantii* to a greater degree than *A. syriaca* due to its more limited ecological and geographic range.

Conversely, gene flow between the two species could potentially be beneficial for the rarer *A. sullivantii*. Hybridization and subsequent introgression can facilitate rapid evolution in response to changing environmental characteristics (Baskett & Gomulkiewicz, 2011). *A. syriaca* is a weedy species and found in a variety of habitats (Bhowmik & Bandeen, 1976). Adaptive gene flow from *A. syriaca* could facilitate population expansion or persistence of *A. sullivantii*, particularly in disturbed habitats. After identifying species specific loci, we could asses shared loci between admixed individuals at different populations, blasting them against the *A. syriaca* reference genome to better understand the molecular profile of gene flow.

If it turns out that the preliminary signature of admixture in this system is not an artefact of our analysis parameters, questions remain about the isolating mechanism in this species pair. As their potential for hybridization and backcrossing has been demonstrated, what is inhibiting hybrids from persisting in co-occurring populations? Klips & Culley showed that flowering time between the species overlaps, and that they share the same generalist pollinators. Based on evidence in other *Asclepias* systems, inability of full pollen tube formation and mechanical isolation due to drastic size differences between stigmatic chamber and pollinia size could be driving hybrid rarity (Hatfield & Kephart, 2003). It could also be possible that only certain genotypes are compatible to form viable hybrids, as parental genotype can influence successful hybridization (Pratt et al., 1985). Further investigations into these mechanisms in our system could yield insight into the barriers to widespread hybridization within this genus.

Our preliminary work confirmed distinct morphological features between parental taxa and hybrid individuals. Early STRUCTURE results show extensive admixture at the population level, but further analysis is required to confirm this. Additional work with this data set has the potential to clarify the degree of genetic exchange between *A. syriaca* and *A. sullivantii*. Taking advantage of our full data set with powerful bioinformatic tools such as HyDe (Blischak et al., 2018), Adegenet (Jombart, 2008), and Arlequin (Excoffier & Lischer, 2010) will allow us to make quantitative inferences about the extent of

admixture in this system, identify species specific genetic markers, and assess population structure within these ecologically important species.

Acknowledgements

Many individuals contributed to this study in a variety of ways, ranging from sample collection, wet lab preparation, and project advice. Thank you to Dr. Steve Hovick for collecting Iowa samples, and supplying constant advice, guidance, reassurance, and last minute edits. Dr. Andreas Chavez provided invaluable advice regarding the genomic analyses and ddRADSeq execution. Thank you to Dr. Bob Klips for providing the inspiration for our study, collecting individuals, and for passing along so much advice and information about *Asclepias* hybridization. Dr. Kattia Palacio-Lopez offered advice and helped with the ddRADSeq preparation. Colin Day collected individuals from populations, extracted hundreds of *Asclepias* DNA samples, and helped with ddRADSeq preparation. Kali Mattingly sampled from Ohio, Wisconsin, and Illinois populations, and has provided constant advice throughout the project. Aaron Tyal assisted in collection of individuals from several Ohio populations. Jimmy Bingman, Mounica Burada, Allie Falla, Austin Kramer, Bailey Madercic, Selinde Roosenburg, and Lizzy Slabinski helped with morphological data collection and DNA quantification on the Nanodrop. Funding from the Mayer's Summer Research Scholarship partially supported the project cost.

References

Allendorf, F. W., Leary, R. F., Spruell, P., & Wenburg, J. K. (2001). The problems with hybrids: setting conservation guidelines. *Trends in Ecology & Evolution*, *16*(11), 613–622. https://doi.org/10.1016/S0169-5347(01)02290-X

Arnold, M. L. (1997). Natural Hybridization and Evolution. Oxford University Press.

Baskett, M. L., & Gomulkiewicz, R. (2011). Introgressive hybridization as a mechanism for species rescue. *Theoretical Ecology*, *4*(2), 223–239. https://doi.org/10.1007/s12080-011-0118-0

Bhowmik, P. C., & Bandeen, J. D. (1976). The Biology of Canadian Weeds: 19. Asclepias syriaca L. *Canadian Journal of Plant Science*, *56*(3), 579–589. https://doi.org/10.4141/cjps76-094

Blischak, P. D., Chifman, J., Wolfe, A. D., & Kubatko, L. S. (2018). HyDe: A Python Package for Genome-Scale Hybridization Detection. *Systematic Biology*, *67*(5), 821–829. https://doi.org/10.1093/sysbio/syy023

Broyles, S. B. (2002). Hybrid Bridges to Gene Flow: A Case Study in Milkweeds (Asclepias). *Evolution*, *56*(10), 1943–1953. JSTOR.

Broyles, S. B., Vail, C., & Sherman-Broyles, S. L. (1996). Pollination genetics of hybridization in sympatric populations of asclepias exaltata and a. syriaca (asclepiadaceae). *American Journal of Botany*, *83*(12), 1580–1584. https://doi.org/10.1002/j.1537-2197.1996.tb12816.x

Cooperrider, T. S. (1997). The dicotyledoneae of Ohio, Part 2. Linaceae through campanulaceae. By T. S. Cooperrider. *Brittonia*, *49*(1), 153–154. https://doi.org/10.2307/2807706

Eaton, D. A. R., & Overcast, I. (2020). ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics*, *36*(8), 2592–2594. https://doi.org/10.1093/bioinformatics/btz966

Ellstrand, N. C. (1992). Gene Flow by Pollen: Implications for Plant Conservation Genetics. *Oikos*, *63*(1), 77–86. https://doi.org/10.2307/3545517

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*(3), 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Fox, J., & Weisberg, S. (2019). *An {R} Companion to Applied Regression* (Third). Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Gleason, H. A., & Cronquist, A. (1991). *Manual of vascular plants of northeastern United States and adjacent Canada* (Vol. 834). New York Botanical Garden.

Graham, C. F., Glenn, T. C., McArthur, A. G., Boreham, D. R., Kieran, T., Lance, S., Manzon, R. G., Martino, J. A., Pierson, T., Rogers, S. M., Wilson, J. Y., & Somers, C. M. (2015). Impacts of

degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). *Molecular Ecology Resources*, *15*(6), 1304–1315. https://doi.org/10.1111/1755-0998.12404

Hatfield, E., & Kephart, S. R. (2003). Reproductive isolation and hybridization between two milkweeds (Asclepias fascicularis and A. speciosa, asclepiadaceae). *Madroño*, *50*(3), 170–180. JSTOR.

John T. Kartesz. (2015). The Biota of North America Program (BONAP). 2015. *Taxonomic Data Center*. (http://www.bonap.net/tdc) Chapel Hill, N.C. [maps generated from Kartesz, J.T. 2015 Floristic Synthesis of North America, Version 1.0. Biota of North America Program (BONAP).

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Kephart, S., & Theiss, K. (2003). Pollinator-mediated isolation in sympatric milkweeds (Asclepias): do floral morphology and insect behavior influence species boundaries? *New Phytologist*, *161*(1), 265–277. https://doi.org/10.1046/j.1469-8137.2003.00956.x

Klips, R. A., & Culley, T. M. (2004). Natural Hybridization between Prairie Milkweeds, Asclepias sullivantii and Asclepias syriaca: Morphological, Isozyme, and Hand-Pollination Evidence. *International Journal of Plant Sciences*, *165*(6), 1027–1037. https://doi.org/10.1086/423875

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, *15*(5), 1179–1191. https://doi.org/10.1111/1755-0998.12387

Mora-Márquez, F., García-Olivares, V., Emerson, B. C., & Heredia, U. L. de. (2017). ddradseqtools: a software package for in silico simulation and testing of double-digest RADseq experiments. *Molecular Ecology Resources*, *17*(2), 230–246. https://doi.org/10.1111/1755-0998.12550

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., & Wagner, H. (2019). *vegan: Community Ecology Package*. https://CRAN.R-project.org/package=vegan

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PLOS ONE*, *7*(5), e37135. https://doi.org/10.1371/journal.pone.0037135

Pratt, R. C., Bressan, R. A., & Hasegawa, P. M. (1985). Genotypic diversity enhances recovery of hybrids and fertile backcrosses of Phaseolus vulgaris L. × P. acutifolius A. Gray. *Euphytica*, *34*(2), 329–334. https://doi.org/10.1007/BF00022926

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, *155*(2), 945–959.

R Wyatt, & Broyles, and S. B. (1994). Ecology and Evolution of Reproduction in Milkweeds. *Annual Review of Ecology and Systematics*, *25*(1), 423–441. https://doi.org/10.1146/annurev.es.25.110194.002231

Ramasamy, R. K., Ramasamy, S., Bindroo, B. B., & Naik, V. G. (2014). STRUCTURE PLOT: a program for drawing elegant STRUCTURE bar plots in user friendly interface. *SpringerPlus*, *3*. https://doi.org/10.1186/2193-1801-3-431

Rhymer, J. M., & Simberloff, D. (1996). Extinction by Hybridization and Introgression. *Annual Review of Ecology and Systematics*, *27*(1), 83–109. https://doi.org/10.1146/annurev.ecolsys.27.1.83

Schloerke, B., Crowley, J., Cook, D., Briatte, F., Marbach, M., Thoen, E., Elberg, A., & Larmarange, J. (2020). *GGally: Extension to "ggplot2"* (Version R package version 1.5.0). https://CRAN.R-project.org/package=GGally

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wickham, H., François, R., Henry, L., & Müller, K. (2019). *dplyr: A Grammar of Data Manipulation. R package version 0.8.3*. https://CRAN.R-project.org/package=dplyr

Woodson, R. E. (1954). The North American Species of Asclepias L. *Annals of the Missouri Botanical Garden*, *41*(1), 1–211. JSTOR. https://doi.org/10.2307/2394652

Supplemental Figures

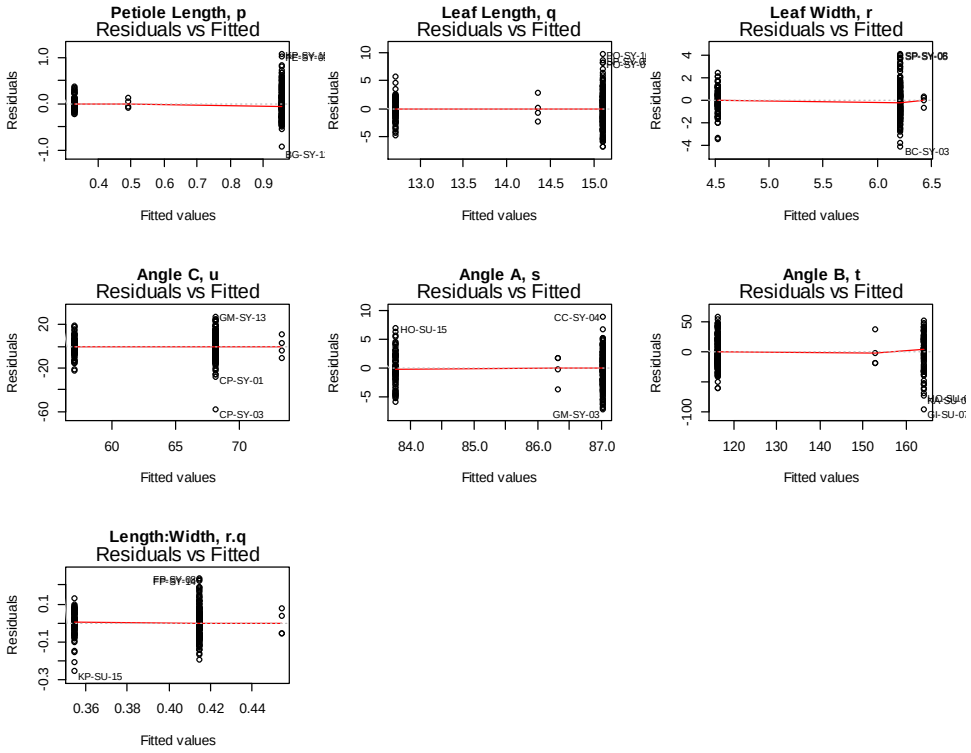

Figure S1: Residual vs. fitted values plots for leaf trait data. Variation in putative hybrid groups is less than that in putative *A. syriaca* and *A. sullivantii*.

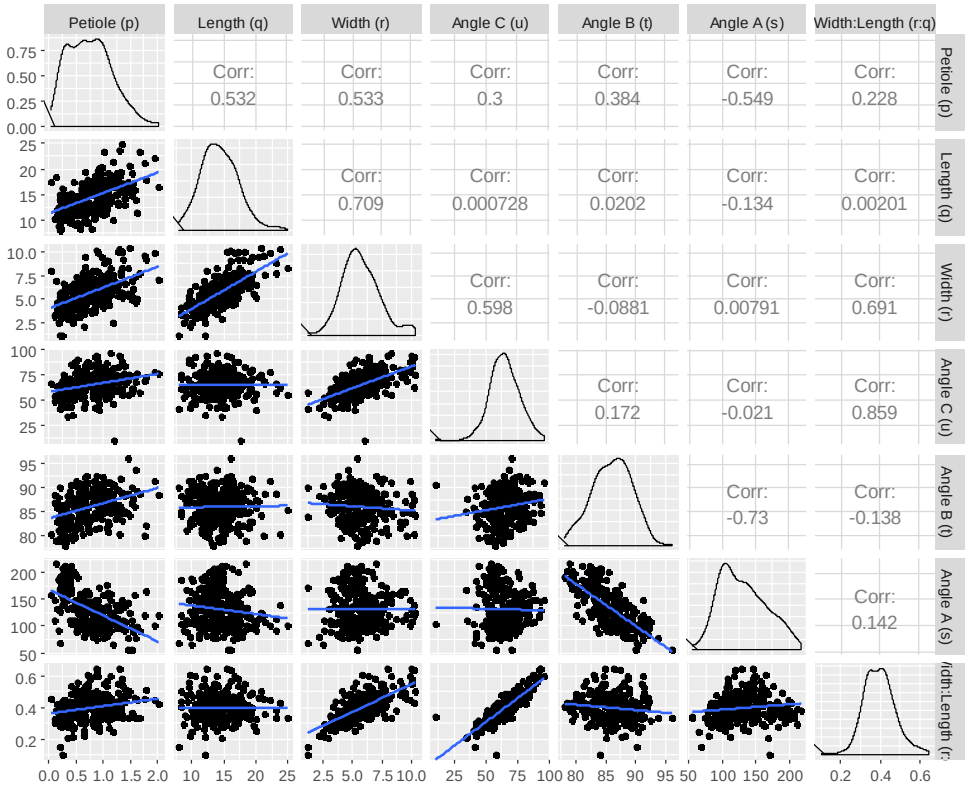

Figure S2: Correlation matrix for leaf trait data.